

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 17:58:38 ; Search time 55 seconds  
(without alignments)  
30.823 Million cell updates/sec

Title: US-09-715-763A-5

Perfect score: 34

Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	6	Aae07240	Peptide #
2	34	100.0	218	Aaw89832	Protein e
3	34	100.0	256	Abp79797	N. gonorr
4	34	100.0	273	Aaw90026	Expressed
5	34	100.0	273	AAW89979	Protein e
6	34	100.0	274	AAW19756	2-Oxoglut
7	34	100.0	326	ABU15067	Protein e
8	31	91.2	14	AAW90308	Human ant
9	31	91.2	122	AAW90284	Human ant
10	31	91.2	210	AAW83113	S. epider
11	31	91.2	210	AAW82885	S. epider
12	31	91.2	288	ABP40124	Staphyloc
13	31	91.2	288	ABW73102	Staphyloc
14	31	91.2	763	AAW20715	Protein e
15	31	91.2	796	AAW70833	C albican
16	31	91.2	921	AAW34522	Prothorm
17	31	91.2	922	AAW34521	Prothorm
18	31	91.2	925	AAW34520	Prothorm
19	31	91.2	938	AAW34392	Prothorm
20	30	88.2	53	AAW18393	Peptide #
21	30	88.2	53	ABB37424	Peptide #
22	30	88.2	53	ABB32184	Peptide #
23	30	88.2	53	ABB22727	Protein #
24	30	88.2	63	ABP31235	Human ORF
25	30	88.2	78	AAW25642	Phleum sp

26	30	88.2	78	7	ADC34891	Timothy g
27	30	88.2	81	7	ADC94212	E. faeciu
28	30	88.2	83	3	AAW19054	Zea mays
29	30	88.2	85	2	AAW25652	Betula sp
30	30	88.2	85	7	ADC34901	Tree alle
31	30	88.2	88	3	AAW06095	Arabidops
32	30	88.2	91	3	AAW44004	Zea mays
33	30	88.2	92	4	AAW43531	Human pol
34	30	88.2	92	4	AAW19932	Novel hum
35	30	88.2	92	4	AAW87507	Novel cen
36	30	88.2	92	4	AAW87196	Novel cen
37	30	88.2	102	4	AAW52864	Propionib
38	30	88.2	102	6	ABW49383	Propionib
39	30	88.2	122	7	ADC96103	Propionib
40	30	88.2	128	3	AAW19053	Zea mays
41	30	88.2	129	3	AAW48315	Arabidops
42	30	88.2	129	3	AAW04159	Arabidops
43	30	88.2	136	3	AAW44003	Zea mays
44	30	88.2	140	2	AAW27559	Cyn d all
45	30	88.2	143	2	AAW27560	Cyn d all

## ALIGNMENTS

## RESULT 1

AAE07240

ID AAE07240 standard; peptide; 6 AA.

XX AC

XX AAE07240;

XX 06-NOV-2001 (first entry)

XX DT

XX DE Peptide #5 used for counteracting effects of ROS and free radicals.

XX KW Antioxidative enzyme; catalase; CAT; superoxide dismutase; SOD; therapy;

XX KW reactive oxygen species; ROS; free radical; dietary supplement; stroke;

XX KW AP-1 transcription factor; renal reperfusion damage; cerebral ischaemia;

XX KW myocardial infarction; heart attack; pain; atherosclerosis; neuroleptic;

XX KW trauma; premature aging; neurodegenerative disease; tardive dyskinesia;

XX KW Parkinson's disease; amyotrophic lateral sclerosis; Alzheimer's disease;

XX KW arthritis; inflammatory disease; diabetes; ulcerative colitis; cataract;

XX KW senility; Down's syndrome; macular degeneration; septic shock; epilepsy;

XX KW polytraumatic shock; schizophrenia; aneurysm; clozapine; tranquilliser;

XX KW cardiant; cerebroprotective; vulnary; nootropic; Huntington's disease;

XX KW anticonvulsant; neuroprotective; antiarthritic; antiinflammatory; burn;

XX KW cytosstatic; leukaemia; ophthalmological; antibacterial;

XX KW immunosuppressive.

XX KW Synthetic.

XX Key

XX Modified-site 1

XX Location/Qualifiers

XX /note= "Optionally protected with glucose-3-O-glycolic

XX acid moiety, palmitoyl group or lipoic acid"

XX WO200136454-A1.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US031764.

XX 18-NOV-1999; 99US-0166381P.

XX (CERE-) CEREMEDIX INC.

XX Shashoua VE;

XX WPI; 2001-496512/54.

XX Novel peptide compound that up regulates expression of a gene encoding

XX antioxiidative enzymes, used to treat or prevent conditions caused by

XX undesirable elevation of reactive oxygen species and other free radicals.

XX PS Claim 4; Page 63; 102pp; English.

XX CC The invention relates to peptide compounds and methods for upregulating

CC expression of a gene encoding an antioxidant enzyme, such as catalase

CC (CAT) or superoxide dismutase (SOD), to counteract harmful oxidative

CC effects of reactive oxygen species (ROS) and other free radicals. The

CC peptides are used as components of pharmaceuticals and dietary

CC supplements. The peptides are used to treat or to prevent diseases and

CC conditions characterised by undesirable elevation of ROS and other free

CC radicals, to upregulate AP-1 transcription factor gene expression and to

CC treat pain. The disease or conditions include renal reperfusion damage,

CC cerebral ischaemia (stroke), myocardial infarction (heart attack), head

CC trauma, atherosclerosis, brain trauma, oxygen toxicity in premature

CC infants, premature aging, spinal cord trauma, neurodegenerative diseases,

CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis,

CC Alzheimer's disease, arthritis and other inflammatory diseases, diabetes,

CC ulcerative colitis, human leukaemia and other cancers characterised by

CC elevation of ROS or free radicals, age-related elevation of ROS or free

CC radicals, senility, Down's syndrome, macular degeneration, cataracts,

CC septic shock, polytraumatic shock, schizophrenia, burn injuries,

CC epilepsy, radiation and/or drug-induced elevation of ROS and free

CC radicals, where the drug is a neuroleptic or a drug such as clozapine

CC defined in the specification and Tardive dyskinesia. The present sequence

CC is a peptide used for counteracting effects of ROS and free radicals

XX SQ Sequence 6 AA;

Query Match 100.0%; Score 34; DB 4; Length 6;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 1 DGDGFA 6

RESULT 2

AAW89832

ID AAW89832 standard; protein; 218 AA.

XX AC AAW89832;

XX AC AAW89832;

DT 20-MAR-2003 (revised)

DT 18-FEB-1999 (first entry)

XX DE Protein encoded by clone a3 ORF2.

XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;

KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX OS Helicobacter pylori.

XX PN WO9849314-A2.

XX PD 05-NOV-1998.

XX PF 25-APR-1998; 98WO-US008487.

XX PF 25-APR-1997; 97US-0045107P.

PR 14-OCT-1997; 97US-0061958P.

XX (GENE-) GENELABS TECHNOLOGIES INC.

PA Chow TP, Fry KE, Lim MY, Meatee CP;

PI WPI; 1999-009433/01.

DR N-PSDB; AAV90766.

XX New Helicobacter pylori antigens and related nucleic acid sequences -

PT useful in serological diagnosis and protective vaccines, providing long-

PT lasting immune response.

XX

PS Claim 15; Page 178; 402pp; English.

XX CC The present sequence represents a Helicobacter pylori antigenic protein

CC that is characterised by immunoreactivity with H. pylori-positive

CC antisera. The proteins are highly immunogenic and induce a long-lasting

CC immune response that persists even after antimicrobial treatment. In

CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are

CC highly sensitive and specific. The specification also describes 69

CC previously unrecognised immunogenic cluster families. H. pylori antigens

CC are used to detect H. pylori-specific antibodies, for diagnosing

CC infection or to confirm eradication of infection, and in vaccines to

CC protect against H. pylori infection and related diseases (gastritis,

CC peptic ulcer, gastric adenocarcinoma/lymphoma). (Updated on 20-MAR-2003

CC to correct PF field.)

XX SQ Sequence 218 AA;

Query Match 100.0%; Score 34; DB 2; Length 218;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 95 DGDGFA 100

RESULT 3

ABP79797

ID ABP79797 standard; protein; 256 AA.

XX AC ABP79797;

XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae amino acid sequence SEQ ID 6124.

XX KW Antibacterial; infection; vaccine; gene therapy.

XX OS Neisseria gonorrhoeae.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB002069.

PR 12-FEB-2001; 2001GB-00003424.

XX (CHIR-) CHIRON SPA.

XX Fontana MR, Pizsa M, Maignani V, Monaci E;

XX WPI; 2003-058415/05.

DR N-PSDB; ABZ40767.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a

XX medicament for treating or preventing N. gonorrhoeae infection.

XX PS Disclosure; Page 635; 815pp; English.

XX CC The present invention relates to proteins from Neisseria gonorrhoeae.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing N. gonorrhoeae

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX SQ Sequence 256 AA;

Query Match 100.0%; Score 34; DB 6; Length 256;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
 |||||  
 Db 49 DGDGFA 54

RESULT 4  
 AAW90026  
 ID AAW90026 standard; protein; 273 AA.  
 AC AAW90026;  
 XX  
 XX 20-MAR-2003 (revised)  
 DT 18-FEB-1999 (first entry)  
 XX  
 XX Expressed antigen for cluster 50.  
 DE  
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;  
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.  
 XX  
 XX Helicobacter pylori.  
 OS  
 XX WO9849314-A2.  
 PN  
 XX 05-NOV-1998.  
 PD  
 XX 25-APR-1998; 98WO-US008487.  
 PF  
 XX 25-APR-1997; 97US-0045107P.  
 PR 14-OCT-1997; 97US-0061958P.  
 XX  
 XX (GENE-) GENELABS TECHNOLOGIES INC.  
 PA  
 XX Chow TP, Fry KE, Lim MY, Mcatee CP;  
 PI WPI; 1999-009433/01.  
 DR N-PSDB; AAV90921.  
 XX  
 XX New Helicobacter pylori antigens and related nucleic acid sequences -  
 PT useful in serological diagnosis and protective vaccines, providing long-  
 PT lasting immune response.  
 XX  
 XX Claim 16; Page 347-348; 402pp; English.  
 PS  
 XX The present sequence represents a Helicobacter pylori antigenic protein  
 CC that is characterised by immunoreactivity with H. pylori-positive  
 CC antisera. The proteins are highly immunogenic and induce a long-lasting  
 CC immune response that persists even after antimicrobial treatment. In  
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are  
 CC highly sensitive and specific. The specification also describes 69  
 CC previously unrecognised immunogenic cluster families. H. pylori antigens  
 CC are used to detect H. pylori-specific antibodies, for diagnosing  
 CC infection or to confirm eradication of infection, and in vaccines to  
 CC protect against H. pylori infection and related diseases (Gastritis,  
 CC peptic ulcer, gastric adenocarcinoma/lymphoma). (Updated on 20-MAR-2003  
 CC to correct PF field.)  
 XX  
 XX Sequence 273 AA;

Query Match 100.0%; Score 34; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 \*Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
 |||||  
 Db 95 DGDGFA 100

RESULT 5  
 AAW89979  
 ID AAW89979 standard; protein; 273 AA.  
 XX  
 AC AAW89979;

XX 20-MAR-2003 (revised)  
 DT 18-FEB-1999 (first entry)  
 XX  
 XX Protein encoded by clone d11 ORF2.  
 DE  
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;  
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.  
 XX  
 XX Helicobacter pylori.  
 OS  
 XX WO9849314-A2.  
 PN  
 XX 05-NOV-1998.  
 PD  
 XX 25-APR-1998; 98WO-US008487.  
 PF  
 XX 25-APR-1997; 97US-0045107P.  
 PR 14-OCT-1997; 97US-0061958P.  
 XX  
 XX (GENE-) GENELABS TECHNOLOGIES INC.  
 PA  
 XX Chow TP, Fry KE, Lim MY, Mcatee CP;  
 PI WPI; 1999-009433/01.  
 DR N-PSDB; AAV90921.  
 XX  
 XX New Helicobacter pylori antigens and related nucleic acid sequences -  
 PT useful in serological diagnosis and protective vaccines, providing long-  
 PT lasting immune response.  
 XX  
 XX Claim 15; Page 319; 402pp; English.  
 PS  
 XX The present sequence represents a Helicobacter pylori antigenic protein  
 CC that is characterised by immunoreactivity with H. pylori-positive  
 CC antisera. The proteins are highly immunogenic and induce a long-lasting  
 CC immune response that persists even after antimicrobial treatment. In  
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are  
 CC highly sensitive and specific. The specification also describes 69  
 CC previously unrecognised immunogenic cluster families. H. pylori antigens  
 CC are used to detect H. pylori-specific antibodies, for diagnosing  
 CC infection or to confirm eradication of infection, and in vaccines to  
 CC protect against H. pylori infection and related diseases (Gastritis,  
 CC peptic ulcer, gastric adenocarcinoma/lymphoma). (Updated on 20-MAR-2003  
 CC to correct PF field.)  
 XX  
 XX Sequence 273 AA;

Query Match 100.0%; Score 34; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
 |||||  
 Db 95 DGDGFA 100

RESULT 6  
 AAW19756  
 ID AAW19756 standard; protein; 274 AA.  
 XX  
 AC AAW19756;  
 XX  
 XX 17-SEP-1997 (first entry)  
 DT  
 XX 2-Oxoglutarate:acceptor oxidoreductase subunit (ORF3 product).  
 DE  
 XX 2-Oxoglutarate:acceptor oxidoreductase; ulcer; vaccine; therapy.  
 KW  
 XX Helicobacter pylori.  
 OS  
 XX WO9723626-A1.  
 XX  
 XX

PD 03-JUL-1997.  
XX  
PF 17-DEC-1996; 96WO-GB0031119.  
XX  
PR 22-DEC-1995; 95GB-00026407.  
XX  
(GLAX ) GLAXO GROUP LTD.  
XX  
PA Chalk PA, Clayton CL, Kelly DJ, Hughes NJ;  
PI  
PL WPI; 1997-351063/32.  
DR N-PSDB; AAT72718.  
XX  
XX New isolated Helicobacter pylori oxido:reductase enzymes - used to  
PT develop products for the diagnosis, treatment and prevention of H. pylori  
PT mediated diseases or disorders.  
XX  
XX Example A; Fig 1; 36pp; English.  
XX  
XX 4 Polypeptides (AAW19754-57) comprise the subunits of a newly identified  
CC tetrameric 2-oxoglutarate:acceptor oxidoreductase (OAO) of Helicobacter  
CC pylori. Their amino acid sequences were deduced from the open reading  
CC frames of a single operon (AAT72718) isolated from a cosmid library. The  
CC ORF3 product has a predicted mol.wt. of 30.6 kDa and shows sequence  
CC similarity to Halobacterium halobium small pyruvate oxidoreductase. The  
CC OAO subunit polypeptides can be expressed in transformed host cells. OAO  
CC and a newly identified pyruvate:ferredoxin oxidoreductase (see also  
CC AAW19758-61) are important or essential to H. pylori but are not  
CC expressed by humans and can therefore be used in vaccines or to screen  
CC for agents that can be used to treat H. pylori diseases or disorders  
XX  
XX Sequence 274 AA;  
SQ  
Query Match 100.0%; Score 34; DB 2; Length 274;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DGDGFA 6  
DB 95 DGDGFA 100  
|||||  
RESULT 7  
ABU15067  
ID ABU15067 standard; protein; 326 AA.  
XX  
AC ABU15067;  
XX  
DT 19-JUN-2003 (first entry)  
DE Protein encoded by Prokaryotic essential gene #594.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Escherichia coli.  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Hagelbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX

DR WPI; 2003-029926/02.  
DR N-PSDB; ACA18937.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 42991; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 326 AA;  
SQ  
Query Match 100.0%; Score 34; DB 6; Length 326;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DGDGFA 6  
DB 78 DGDGFA 83  
|||||  
RESULT 8  
AAW90308  
ID AAW90308 standard; protein; 14 AA.  
XX  
XX AAW90308;  
XX  
XX 07-SEP-1999 (first entry)  
XX  
XX Human anti-idiotypic antibody heavy chain protein CDR3 region 4.  
XX  
XX Antibody; GPr1b/IIa; human; auto-antibody; anti-idiotypic; diagnosis;  
XX blood platelet membrane protein; predisposition; prevention; treatment;  
XX autoimmune thrombocytopenic purpura; AITP; fibrinogen binding; thrombi;  
XX thrombocyte; cardiac infarction; pulmonary embolism; heavy chain; CDR3.  
XX  
XX Homo sapiens.  
XX  
XX WO9855619-A1.  
XX  
XX 10-DEC-1998.  
XX

PF 05-JUN-1998; 98WO-EP003397.  
 XX  
 PR 06-JUN-1997; 97DE-01023904.  
 PR 12-DEC-1997; 97DE-01055227.  
 PR 08-MAY-1998; 98DE-01020663.  
 XX  
 PA (ASAT-) ASAT AG APPLIED SCI & TECHNOLOGY.  
 XX  
 PI Berchtold P, Escher RFA;  
 XX  
 XX WPI; 1999-105496/09.  
 XX  
 DR Nucleic acid encoding human autoantibodies against platelet glycoprotein  
 PT IIB/IIIA - used for diagnosis, treatment and prevention of autoimmune  
 PT thrombocytopaenic purpura and for modulation of fibrinogen binding.  
 XX  
 PS Claim 7; Page 7; 93pp; German.  
 XX  
 CC This invention describes novel nucleic acid fragments that encode human  
 CC auto-antibodies and anti-idiotypic antibodies against blood platelet  
 CC membrane protein, GPIIb/IIIa. The products of the invention are used for  
 CC diagnosis (including monitoring and determining predisposition),  
 CC prevention and treatment of autoimmune thrombocytopaenic purpura (AITP)  
 CC and also for modulating binding of fibrinogen to thrombocytes  
 CC (particularly to dissolve thrombi and/or prevent their formation, e.g. in  
 CC cases of cardiac infarction or pulmonary embolism). Unlike murine  
 CC antibodies, human antibodies (hAb) do not induce adverse side effects and  
 CC persist for longer in vivo than small peptides. AAW90293-W90337 represent  
 CC antibody fragments used in the method of the invention  
 XX  
 SQ Sequence 14 AA;  
 XX  
 Query Match 91.2%; Score 31; DB 2; Length 14;  
 Best Local Similarity 83.3%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 Db 3 DGDGFS 8  
 RESULT 9  
 AAW90284  
 ID AAW90284 standard; protein; 122 AA.  
 XX  
 AC AAW90284;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE Human anti-GPIIb/IIIa antibody heavy chain protein from phagemid A1-X20.  
 XX  
 KW Antibody; GPIIb/IIIa; human; auto-antibody; anti-idiotypic; diagnosis;  
 KW blood platelet membrane protein; predisposition; prevention; treatment;  
 KW autoimmune thrombocytopaenic purpura; AITP; fibrinogen binding; thrombi;  
 KW thrombocyte; cardiac infarction; pulmonary embolism; heavy chain.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 FH 1. .30  
 FT Region /label= FR1  
 FT /note= "framework region 1"  
 FT 31. .35  
 FT Region /label= CDR1  
 FT /note= "complementarity determining region 1"  
 FT 36. .49  
 FT Region /label= FR2  
 FT /note= "framework region 2"  
 FT 50. .65  
 FT Region /label= CDR2  
 FT /note= "complementarity determining region 2"  
 FT 66. .97  
 FT Region /label= FR3

FT /note= "framework region 3"  
 FT 98. .111  
 FT /label= CDR3  
 FT /note= "complementarity determining region 3"  
 FT 112. .122  
 FT /label= FR4  
 FT /note= "framework region 4"  
 XX  
 PN WO9855619-A1.  
 XX  
 PD 10-DEC-1998.  
 XX  
 XX 05-JUN-1998; 98WO-EP003397.  
 XX  
 PR 06-JUN-1997; 97DE-01023904.  
 PR 12-DEC-1997; 97DE-01055227.  
 PR 08-MAY-1998; 98DE-01020663.  
 XX  
 PA (ASAT-) ASAT AG APPLIED SCI & TECHNOLOGY.  
 XX  
 PI Berchtold P, Escher RFA;  
 XX  
 XX WPI; 1999-105496/09.  
 XX N-PSDB; AAV72229.  
 XX  
 DR Nucleic acid encoding human autoantibodies against platelet glycoprotein  
 PT IIB/IIIA - used for diagnosis, treatment and prevention of autoimmune  
 PT thrombocytopaenic purpura and for modulation of fibrinogen binding.  
 XX  
 PS Disclosure; Page 56; 93pp; German.  
 XX  
 CC This invention describes novel nucleic acid fragments that encode human  
 CC auto-antibodies and anti-idiotypic antibodies against blood platelet  
 CC membrane protein, GPIIb/IIIa. The products of the invention are used for  
 CC diagnosis (including monitoring and determining predisposition),  
 CC prevention and treatment of autoimmune thrombocytopaenic purpura (AITP)  
 CC and also for modulating binding of fibrinogen to thrombocytes  
 CC (particularly to dissolve thrombi and/or prevent their formation, e.g. in  
 CC cases of cardiac infarction or pulmonary embolism). Unlike murine  
 CC antibodies, human antibodies (hAb) do not induce adverse side effects and  
 CC persist for longer in vivo than small peptides  
 XX  
 SQ Sequence 122 AA;  
 XX  
 Query Match 91.2%; Score 31; DB 2; Length 122;  
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 Db 100 DGDGFS 105  
 RESULT 10  
 AAG83113  
 ID AAG83113 standard; protein; 210 AA.  
 XX  
 AC AAG83113;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:3320.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;  
 KW endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US030782.

Wed Jun 2 16:48:16 2004

us-09-715-763a-5.rag

```
XX 09-NOV-1999; 99US-0164258P.
PR (GLAX ) GLAXO GROUP LTD.
PA Kimmerly WJ;
PI
XX WPI; 2001-316495/33.
DR N-PSDB; AAH53963.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX Claim 18; Page 875; 2189pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AG81454 to AG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX Sequence 210 AA;
SQ
Query Match 91.2%; Score 31; DB 4; Length 210;
Best Local Similarity 83.3%; Pred. NO. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGDGFA 6
Db 14 DGDGYA 19
|||||
RESULT 11
AAG82885
XX AAG82885 standard; protein; 210 AA.
XX AAG82885;
XX 03-SEP-2001 (first entry)
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2864.
XX Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX endocarditis.
XX Staphylococcus epidermidis.
XX WO200134809-A2.
XX 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US030782.
XX 09-NOV-1999; 99US-0164258P.
XX (GLAX ) GLAXO GROUP LTD.
XX Kimmerly WJ;
PI
XX 09-NOV-1999; 99US-0164258P.
PR (GLAX ) GLAXO GROUP LTD.
PA Kimmerly WJ;
PI
```

---

```
XX WPI; 2001-316495/33.
DR N-PSDB; AAH53735.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX Claim 18; Page 747; 2189pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AG81454 to AG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX Sequence 210 AA;
SQ
Query Match 91.2%; Score 31; DB 4; Length 210;
Best Local Similarity 83.3%; Pred. NO. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGDGFA 6
Db 14 DGDGYA 19
|||||
RESULT 12
ABP40124
ID ABP40124 standard; protein; 288 AA.
XX ABP40124;
XX 24-JUL-2002 (first entry)
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4969.
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX Staphylococcus epidermidis.
XX US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-00134001.
XX 14-AUG-1997; 97US-0055779P.
XX 08-NOV-1997; 97US-0064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92669.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT
```

PT polypeptide, useful for diagnosing and treating bacterial infections.  
 PS Disclosure; SEQ ID NO 4969; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 288 AA;

Query Match 91.2%; Score 31; DB 5; Length 288;  
 Best Local Similarity 83.3%; Pred. No. 5.9e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 Db 92 DGDGYA 97  
 |||||:  
 |||||:

RESULT 13  
 ABW73102  
 ID ABW73102 standard; protein; 288 AA.  
 XX  
 AC ABW73102;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE *Staphylococcus aureus* protein #2342.  
 XX  
 KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
 KW enzymatic assay; antibiotic target.  
 XX  
 OS *Staphylococcus aureus*.  
 XX  
 XX WO200294868-A2.  
 PN  
 XX 28-NOV-2002.  
 PD  
 XX 27-MAR-2002; 2002WO-IB002637.  
 PF  
 XX 27-MAR-2001; 2001GB-00007661.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA  
 PI Masignani V, Mora M, Scarselli M;  
 PI WPI; 2003-120786/11.  
 DR N-PSDB; ACF74662.  
 DR  
 XX New *Staphylococcus aureus* protein, useful as a vaccine for treating or  
 PT preventing *Staphylococcal* infection, specifically an infection caused by  
 PT *S. aureus*, e.g. sepsis.  
 XX  
 XX Claim 1; SEQ ID NO 4684; 49pp; English.  
 PS  
 XX The invention relates to novel genes and encoded proteins from  
 CC *Staphylococcus aureus*. A composition comprising the *S. aureus* protein, a  
 CC nucleic acid encoding the protein, or an antibody to the protein, is  
 CC useful as a pharmaceutical, particularly as a vaccine for treating or  
 CC preventing infection due to *Staphylococcus aureus*, specifically an  
 CC infection caused by *S. aureus*. The composition is particularly useful for  
 CC treating or preventing sepsis in a patient. The composition can also be  
 CC used for diagnostics. The protein is also used in an assay for enzymatic  
 CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel *S. aureus* proteins of the invention  
 XX

SQ Sequence 288 AA;

Query Match 91.2%; Score 31; DB 6; Length 288;  
 Best Local Similarity 83.3%; Pred. No. 5.9e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 Db 92 DGDGYA 97  
 |||||:  
 |||||:

RESULT 14  
 ABU20715  
 ID ABU20715 standard; protein; 763 AA.  
 XX  
 AC ABU20715;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by *Prokaryotic essential gene #6242*.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Bacteroides fragilis.  
 OS  
 XX WO200277183-A2.  
 PN  
 XX 03-OCT-2002.  
 PD  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 PI WPI; 2003-029926/02.  
 DR N-PSDB; ACA24585.  
 DR  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 48639; 176pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for

us-09-715-763a-5.rag

Wed Jun 2 16:48:16 2004

QY 1 DGDGFA 6  
DB 78 DGDGFA 83  
Search completed: June 1, 2004, 18:08:38  
Job time : 57 secs

CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 763 AA;

Query Match 91.2%; Score 31; DB 6; Length 763;  
Best Local Similarity 83.3%; Pred. No. 1.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
DB 292 DGDGFS 297

RESULT 15  
AAG70833  
ID AAG70833 standard; protein; 796 AA.

XX AC AAG70833;  
XX DT 27-JUL-2001 (first entry)  
XX DE C albicans apoptosis associated protein #13.  
XX DE Yeast; fungus; apoptosis; infection; proliferative disease; vaccine;  
XX KW autoimmune disease; ischaemia; neurodegeneration.  
XX OS Candida albicans.

XX PN WO200102550-A2.  
XX PD 11-JAN-2001.  
XX PF 03-JUL-2000; 2000WO-BE0000077.  
XX PR 01-JUL-1999; 99EP-00870141.  
XX PA (JANC ) JANSSEN PHARM NV.

XX PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;  
XX PI Nelissen BJM, Reekmans RJ;  
XX DR WPI; 2001-367042/38.  
XX DR N-PSDB; AAH29869.  
XX PT Yeast and fungal nucleic acids encoding proteins involved in a pathway  
XX PT leading to programmed cell death, useful for treating proliferative  
XX PT disorders, yeast and fungal infections, or for preventing apoptosis in  
XX PT certain diseases.

XX PS Claim 24; Fig 2; 218pp; English.  
XX CC The present invention provides the protein and coding sequences of a  
XX CC number of apoptosis associated proteins from the yeast *Saccharomyces*  
XX CC *cerevisiae* and the fungus *Candida albicans*. These can be used to identify  
XX CC treatments for fungal and yeast infections, for proliferative diseases  
XX CC and for apoptosis related diseases such as autoimmune diseases, ischaemia  
XX CC and neurodegeneration. The present sequence is one of the *C. albicans*  
XX CC proteins of the invention  
XX

SQ Sequence 796 AA;  
Query Match 91.2%; Score 31; DB 4; Length 796;  
Best Local Similarity 83.3%; Pred. No. 1.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 18:07:34 ; Search time 23 Seconds  
(without alignments)  
13.468 Million cell updates/sec

Title: US-09-715-763A-5  
Perfect score: 34  
Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	91.2	268	4	US-09-252-991A-21630
2	31	91.2	288	4	US-09-134-001C-4969
3	30	88.2	81	4	US-09-107-532A-3839
4	30	88.2	116	4	US-09-489-039A-7242
5	30	88.2	122	4	US-09-107-532A-5730
6	30	88.2	219	4	US-09-205-258-271
7	30	88.2	264	4	US-09-252-991A-19619
8	30	88.2	291	4	US-09-252-991A-25517
9	30	88.2	331	2	US-08-828-242-3
10	30	88.2	331	2	US-08-910-927B-5
11	30	88.2	331	3	US-09-206-499-3
12	30	88.2	331	3	US-09-270-270-5
13	30	88.2	356	4	US-09-540-236-3392
14	30	88.2	367	4	US-09-404-296B-6
15	30	88.2	381	4	US-09-328-352-4500
16	30	88.2	478	4	US-09-489-039A-12483
17	30	88.2	496	4	US-09-312-762A-15
18	30	88.2	520	4	US-09-252-991A-17942
19	30	88.2	529	1	US-08-178-477B-32
20	30	88.2	529	4	US-09-304-121-2
21	30	88.2	573	4	US-09-252-991A-31334
22	30	88.2	748	4	US-09-134-001C-6041
23	30	88.2	782	4	US-09-360-545-16
24	30	88.2	782	4	US-09-398-395A-46
25	30	88.2	782	4	US-09-887-586A-46
26	30	88.2	782	4	US-09-895-752-46
27	30	88.2	782	4	US-09-903-012B-46

28	30	88.2	782	4	US-09-900-797-46	Sequence 46, Appl
29	30	88.2	783	4	US-09-513-783A-176	Sequence 176, App
30	30	88.2	817	3	US-09-234-393-13	Sequence 13, Appl
31	30	88.2	817	3	US-09-234-393-38	Sequence 38, Appl
32	30	88.2	817	3	US-09-234-393-40	Sequence 40, Appl
33	30	88.2	817	3	US-09-234-393-42	Sequence 42, Appl
34	30	88.2	817	4	US-09-865-171-13	Sequence 13, Appl
35	30	88.2	817	4	US-09-865-171-38	Sequence 38, Appl
36	30	88.2	817	4	US-09-865-171-40	Sequence 40, Appl
37	30	88.2	817	4	US-09-865-171-42	Sequence 42, Appl
38	30	88.2	829	4	US-09-252-991A-28854	Sequence 28854, A
39	30	88.2	932	4	US-09-071-035-416	Sequence 416, App
40	30	88.2	969	4	US-09-071-035-414	Sequence 414, App
41	30	88.2	1007	4	US-09-252-991A-28773	Sequence 28773, A
42	30	88.2	1034	4	US-09-562-737-82	Sequence 82, Appl
43	29	85.3	139	4	US-09-489-039A-12006	Sequence 12006, A
44	29	85.3	476	4	US-09-489-039A-13646	Sequence 13646, A
45	29	85.3	1065	4	US-09-328-352-5536	Sequence 5536, Ap

## ALIGNMENTS

## RESULT 1

US-09-252-991A-21630  
; Sequence 21630, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21630

; LENGTH: 268

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21630

Query Match Best Local Similarity 91.2%; Score 31; DB 4; Length 268;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6

Db 37 DGDGFS 42

## RESULT 2

US-09-134-001C-4969

; Sequence 4969, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GFC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4969

; LENGTH: 288

; TYPE: PRT

us-09-715-763a-5.ra1

Wed Jun 2 16:48:16 2004

```
US-09-489-039A-7242
; Sequence 7242, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7242
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7242

Query Match      88.2%; Score 30; DB 4; Length 116;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGDGF 5
DB      51 DGDGF 55

RESULT 5
US-09-107-532A-5730
; Sequence 5730, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-5007
; INFORMATION FOR SEQ ID NO: 5730:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 81 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...81
; SEQUENCE DESCRIPTION: SEQ ID NO: 3839:
US-09-107-532A-3839

Query Match      88.2%; Score 30; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGDGF 5
DB      50 DGDGF 54

RESULT 4
US-09-134-001C-4969
; ORGANISM: Staphylococcus epidermidis
; Query Match      91.2%; Score 31; DB 4; Length 288;
; Best Local Similarity 83.3%; Pred. No. 1.7e+02;
; Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGDGFA 6
DB      92 DGDGYA 97

RESULT 3
US-09-107-532A-3839
; Sequence 3839, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-5007
; INFORMATION FOR SEQ ID NO: 3839:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 81 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...81
; SEQUENCE DESCRIPTION: SEQ ID NO: 3839:
US-09-107-532A-3839

Query Match      88.2%; Score 30; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGDGF 5
DB      50 DGDGF 54
```

FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...122  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5730:  
 US-09-107-532A-5730

Query Match 88.2%; Score 30; DB 4; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGF 5  
 Db 66 DGDGF 70

## RESULT 6

US-09-205-258-271  
 ; Sequence 271, Application US/09205258  
 ; Patent No. 6525174  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young et al.  
 ; TITLE OF INVENTION: 207 Human Secreted Proteins  
 ; FILE REFERENCE: P2007P1  
 ; CURRENT APPLICATION NUMBER: US/09/205,258  
 ; CURRENT FILING DATE: 1998-12-04  
 ; EARLIER APPLICATION NUMBER: PCT/US98/11422  
 ; EARLIER FILING DATE: 1998-06-04  
 ; EARLIER APPLICATION NUMBER: 60/048,885  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/049,375  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,881  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,880  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,896  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/049,020  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,876  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,895  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,884  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,894  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,971  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,964  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,882  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,899  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,893  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,900  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,901  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,892  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,915  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/049,019  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,970  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,972  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,916  
 ; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,373  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,875  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/049,374  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,917  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,949  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,974  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,883  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,897  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,898  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,962  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,963  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,877  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/048,878  
 ; EARLIER FILING DATE: 1997-06-06  
 ; EARLIER APPLICATION NUMBER: 60/070,923  
 ; EARLIER FILING DATE: 1997-12-18  
 ; EARLIER APPLICATION NUMBER: 60/092,921  
 ; EARLIER FILING DATE: 1998-07-15  
 ; EARLIER APPLICATION NUMBER: 60/094,657  
 ; EARLIER FILING DATE: 1998-07-30  
 ; NUMBER OF SEQ ID NOS: 1227  
 ; SOFTWARE: Patentin ver. 2.0  
 ; SEQ ID NO 271  
 ; LENGTH: 219  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (219)  
 ; OTHER INFORMATION: Xaa equals stop translation  
 US-09-205-258-271

Query Match 88.2%; Score 30; DB 4; Length 219;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGF 5  
 Db 200 DGDGF 204

## RESULT 7

US-09-252-991A-19619  
 ; Sequence 19619, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 19619  
 ; LENGTH: 264  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-19619

Query Match  
Best Local Similarity 88.2%; Score 30; DB 4; Length 264;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
Db 98 DGDGF 102

## RESULT 8

US-09-252-991A-25517  
Sequence 25517, Application US/09252991A  
Patent No. 6551795

## GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25517  
LENGTH: 291

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25517

Query Match  
Best Local Similarity 88.2%; Score 30; DB 4; Length 291;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
Db 168 DGDGF 172

## RESULT 9

US-08-828-242-3  
Sequence 3, Application US/08828242  
Patent No. 5871970

## GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL CALCIUM-BINDING  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/828,242  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0261 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 331 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1262329  
US-08-828-242-3

Query Match  
Best Local Similarity 88.2%; Score 30; DB 2; Length 331;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
Db 94 DGDGF 98

## RESULT 10

US-08-910-927B-5  
Sequence 5, Application US/08910927B  
Patent No. 5976801

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: HUMAN RETICULOCALBIN ISOFORMS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,927B  
FILING DATE: Hereewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0358 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 331 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1262329

US-08-910-927B-5

Query Match 88.2%; Score 30; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGF 5  
Db 94 DGDGF 98

RESULT 11

US-09-206-499-3  
; Sequence 3, Application US/09206499  
; Patent No. 6194385

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Goli, Surya K. NOVEL CALCIUM-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/206,499

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/828,242

; FILING DATE: 03/31/1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0261 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 331 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 1262329

US-09-206-499-3

Query Match 88.2%; Score 30; DB 3; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGF 5  
Db 94 DGDGF 98

RESULT 12

US-09-270-270-5

; Sequence 5, Application US/09270270

; Patent No. 6235477

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN RETICULOCALBIN ISOFORMS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/270,270

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/910,927

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0358 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 331 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 1262329

US-09-270-270-5

Query Match 88.2%; Score 30; DB 3; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGF 5  
Db 94 DGDGF 98

RESULT 13

US-09-540-236-3392

; Sequence 3392, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRH

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 3392

; LENGTH: 356

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-3392

Query Match 88.2%; Score 30; DB 4; Length 356;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
| | | |  
Db 225 DGDGF 229

## RESULT 14

US-09-404-296B-6  
; Sequence 6, Application US/09404296B  
; Patent No. 6559358  
; GENERAL INFORMATION:  
; APPLICANT: MURRAY, James Augustus Henry  
; TITLE OF INVENTION: PLANTS WITH MODIFIED GROWTH  
; FILE REFERENCE: 2121-0151P  
; CURRENT APPLICATION NUMBER: US/09/404,296B  
; CURRENT FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Nicotiana tabacum  
US-09-404-296B-6

Query Match 88.2%; Score 30; DB 4; Length 367;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
| | | |  
Db 82 DGDGF 86

## RESULT 15

US-09-328-352-4500  
; Sequence 4500, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4500  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4500

Query Match 88.2%; Score 30; DB 4; Length 381;

Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
| | | |  
Db 40 DGDGF 44

Search completed: June 1, 2004, 18:11:02

Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 18:08:44 ; Search time 43 Seconds  
(without alignments)  
38.965 Million cell updates/sec

Title: US-09-715-763A-5  
Perfect score: 34  
Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	237	12	US-10-335-977-6441
2	34	100.0	238	12	Sequence 6441, Ap
3	34	100.0	260	12	Sequence 6440, Ap
4	34	100.0	273	12	US-10-335-977-6442
5	34	100.0	326	12	Sequence 6442, Ap
6	34	100.0	523	12	US-10-335-977-6443
7	31	91.2	475	14	Sequence 6443, Ap
8	31	91.2	481	14	Sequence 240458, A
9	31	91.2	489	14	US-10-424-599-240458
10	31	91.2	763	12	Sequence 12821, A
11	30	88.2	53	9	US-10-156-761-12821
12	30	88.2	63	11	Sequence 12010, A
13	30	88.2	65	12	Sequence 8688, Ap
14	30	88.2	70	14	US-10-156-761-12010
15	30	88.2	76	12	Sequence 416, App
					Sequence 271030, A
					Sequence 10317, A
					Sequence 255937, A

16 30 88.2 81 12 US-10-424-599-162802  
17 30 88.2 88 12 US-10-424-599-204845  
18 30 88.2 92 10 US-09-764-881-129  
19 30 88.2 92 11 US-09-764-875-714  
20 30 88.2 92 11 US-09-764-875-1025  
21 30 88.2 92 12 US-09-764-881-129  
22 30 88.2 92 15 US-10-242-747-129  
23 30 88.2 92 15 US-10-158-057-209  
24 30 88.2 122 12 US-10-424-599-205039  
25 30 88.2 141 12 US-10-424-599-238414  
26 30 88.2 141 12 US-10-424-599-263508  
27 30 88.2 141 14 US-10-156-761-7754  
28 30 88.2 156 12 US-10-424-599-184133  
29 30 88.2 187 12 US-10-425-114-67716  
30 30 88.2 189 12 US-10-424-599-147725  
31 30 88.2 219 10 US-09-933-767-271  
32 30 88.2 219 12 US-10-004-860-271  
33 30 88.2 219 14 US-10-023-282-271  
34 30 88.2 222 10 US-09-946-374-180  
35 30 88.2 222 12 US-10-006-485A-180  
36 30 88.2 222 12 US-10-013-907A-180  
37 30 88.2 222 12 US-10-015-499A-180  
38 30 88.2 222 12 US-10-013-910A-180  
39 30 88.2 222 12 US-10-226-254A-180  
40 30 88.2 222 12 US-10-015-395A-180  
41 30 88.2 222 14 US-10-006-856A-180  
42 30 88.2 222 14 US-10-006-818A-180  
43 30 88.2 222 14 US-10-015-393A-180  
44 30 88.2 222 14 US-10-015-869A-180  
45 30 88.2 222 14 US-10-012-121A-180

## ALIGNMENTS

RESULT 1  
US-10-335-977-6441  
; Sequence 6441, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 6441:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 237 amino acids

us-09-715-763a-5.rapb

Wed Jun 2 16:48:16 2004

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori

## FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...237  
SEQUENCE DESCRIPTION: SEQ ID NO: 6441:

US-10-335-977-6441

Query Match 100.0%; Score 34; DB 12; Length 237;  
Best Local Similarity 100.0%; Pred. No. 1e+02; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

DB 81 DGDGFA 86

## RESULT 2

US-10-335-977-6440  
; Sequence 6440, Application US/10335977  
; Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 6440:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

## FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...238  
SEQUENCE DESCRIPTION: SEQ ID NO: 6440:

US-10-335-977-6440

Query Match 100.0%; Score 34; DB 12; Length 238;  
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
DB 82 DGDGFA 87

## RESULT 3

US-10-335-977-6442  
; Sequence 6442, Application US/10335977  
; Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 6442:

SEQUENCE CHARACTERISTICS:

LENGTH: 260 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

## FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...260  
SEQUENCE DESCRIPTION: SEQ ID NO: 6442:

US-10-335-977-6442

Query Match 100.0%; Score 34; DB 12; Length 260;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

DB 82 DGDGFA 87

## RESULT 4

US-10-335-977-6443  
; Sequence 6443, Application US/10335977  
; Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES



RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6443:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...273  
SEQUENCE DESCRIPTION: SEQ ID NO: 6443:  
US-10-335-977-6443

Query Match 100.0%; Score 34; DB 12; Length 273;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
DB 95 DGDGFA 100

RESULT 5  
US-10-282-122A-42991  
Sequence 42991, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 42991  
LENGTH: 326  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-282-122A-42991

Query Match 100.0%; Score 34; DB 12; Length 326;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
DB 78 DGDGFA 83

RESULT 6  
US-10-424-599-240458  
Sequence 240458, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 240458  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)...(523)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_59160C.1.pep  
US-10-424-599-240458

Query Match 100.0%; Score 34; DB 12; Length 523;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
DB 365 DGDGFA 370

Wed Jun 2 16:48:16 2004

us-09-715-763a-5.rapb

```
RESULT 7
US-10-156-761-12821
; Sequence 12821, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: ISHIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12821
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12821

Query Match          91.2%; Score 31; DB 14; Length 475;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DGDGFA 6
Db      121 DGDGFS 126

RESULT 8
US-10-156-761-12010
; Sequence 12010, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: ISHIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12010
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12010

Query Match          91.2%; Score 31; DB 14; Length 481;
Best Local Similarity 83.3%; Pred. No. 7.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DGDGFA 6
Db      106 DGDGYA 111

RESULT 9
US-10-156-761-8688
; Sequence 8688, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIYAKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8688
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8688

Query Match          91.2%; Score 31; DB 14; Length 489;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DGDGFA 6
Db      360 DGDGYA 365

RESULT 10
US-10-282-122A-48639
; Sequence 48639, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
```

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48639
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48639

Query Match      91.2%; Score 31; DB 12; Length 763;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 292 DGDGFS 297

RESULT 11
US-09-864-761-38025
; Sequence 38025, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38025
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009948.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: Q62703, EVALUE 6.80e-01
; OTHER INFORMATION: EST_HUMAN HIT: A127150.1, EVALUE 8.00e-26
US-09-864-761-38025
```

```
Query Match      88.2%; Score 30; DB 9; Length 53;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGDGF 5
Db 35 DGDGF 39
```

## RESULT 12

```
US-09-864-408A-416
; Sequence 416, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 416
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-416
```

```
Query Match      88.2%; Score 30; DB 11; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGDGF 5
Db 11 DGDGF 15
```

## RESULT 13

```
US-10-424-599-271030
; Sequence 271030, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271030
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Glycine max
```

us-09-715-763a-5.rapb

Wed Jun 2 16:48:16 2004

```
; NAME/KEY: unsure
; LOCATION: (1)..(76)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_73131C.1.pap
US-10-424-599-255937

Query Match      88.2%; Score 30; DB 12; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5
Db 48 DGDGF 52

Search completed: June 1, 2004, 18:11:58
Job time : 44 secs
```

```
; NAME/KEY: unsure
; LOCATION: (1)..(65)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_86759C.1.pap
US-10-424-599-271030

Query Match      88.2%; Score 30; DB 12; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5
Db 55 DGDGF 59
```

```
RESULT 14
US-10-156-761-10317
; Sequence 10317, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10317
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10317

Query Match      88.2%; Score 30; DB 14; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5
Db 17 DGDGF 21
```

```
RESULT 15
US-10-424-599-255937
; Sequence 255937, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 255937
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 18:06:04 ; Search time 21 Seconds  
(without alignments)  
27.483 Million cell updates/sec

Title: US-09-715-763A-5

Perfect score: 34

Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	273	1 F64593	pyruvate synthase
2	34	100.0	273	2 H71919	chain of 2-oxoglutarate
3	34	100.0	288	2 E83946	pyruvate synthase
4	34	100.0	310	2 C95950	conserved hypothet
5	34	100.0	310	2 T35417	probable beta-lact
6	34	100.0	314	2 AC2878	aromatic compounds
7	34	100.0	325	2 E97654	ABC transporter ho
8	34	100.0	326	2 B91100	hypothetical prote
9	34	100.0	326	2 B65074	hypothetical prote
10	34	100.0	326	2 D84325	hypothetical prote
11	34	100.0	536	2 D84325	Htr17 transducer [
12	34	100.0	627	2 F84194	Htr14 transducer [
13	34	100.0	764	2 A84328	Htr2 transducer [i
14	34	100.0	765	1 T44946	transducer protein
15	34	100.0	777	2 T44897	transducer protein
16	34	100.0	778	2 T44897	Htr4 transducer [i
17	34	100.0	778	2 F84237	transducer protein
18	34	100.0	788	2 T44262	transducer protein
19	34	100.0	789	2 E84236	transducer protein
20	31	91.2	262	2 T17495	Htr6 transducer [i
21	31	91.2	267	2 C83242	ankyrin repeat pro
22	31	91.2	288	2 E89903	conserved hypothet
23	31	91.2	339	2 T28138	hypothetical prote
24	31	91.2	340	2 T28137	Ig V-region-like B
25	31	91.2	372	2 C39371	Ig V-region-like B
26	31	91.2	536	1 A47190	Ig V-region-like B
27	31	91.2	536	2 E84318	transducer protein
28	31	91.2	1025	2 AE2243	Htr1 transducer [i
29	31	91.2	1408	2 H69068	hypothetical prote cell surface glyco

30	30	88.2	83	2	S65144	pollen allergen gr
31	30	88.2	83	2	S65145	pollen allergen gr
32	30	88.2	85	2	S54819	BETV4 protein - Eu
33	30	88.2	85	2	JC5711	pollen allergen Be
34	30	88.2	146	2	T51473	calmodulin-like pr
35	30	88.2	147	1	MCJ2R	calmodulin - sea p
36	30	88.2	148	1	MCSW	calmodulin - scall
37	30	88.2	148	1	MCXAM	calmodulin - sea a
38	30	88.2	151	2	T49071	probable calmoduli
39	30	88.2	153	2	E88955	protein K04F1.8 [i
40	30	88.2	169	2	D84864	probable calcium b
41	30	88.2	187	2	A84532	probable calmoduli
42	30	88.2	189	2	D84289	hypothetical prote
43	30	88.2	195	2	T49012	calmodulin-like pr
44	30	88.2	220	2	B84281	riboflavin-specifi
45	30	88.2	231	2	E70911	probable partial C

## ALIGNMENTS

### RESULT 1

F64593

pyruvate synthase (EC 1.2.7.1) beta chain - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 05-May-2000

C:Accession: F64593

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: F64593

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <TOM>

A:Cross-references: GB:AE000572; GB:AE000511; NID:g2313703; PIDN:AAD07655.1; PID:g2313705

C:Superfamily: pyruvate synthase beta chain

C:Keywords: coenzyme A; oxidoreductase

Query Match Best Local Similarity 100.0%; Score 34; DB 1; Length 273;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 95 DGDGFA 100

### RESULT 2

H71919

chain of 2-oxoglutarate oxidoreductase - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 11-Jun-1999

C:Accession: H71919

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:9923692

A:Accession: H71919

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <ARN>

A:Cross-references: GB:AE001486; GB:AE001439; NID:g4155074; PIDN:AAD06120.1; PID:g415508

A:Experimental source: strain J99

C:Genetics:

A:Gene: oorB

C:Superfamily: pyruvate synthase beta chain

Query Match 100.0%; Score 34; DB 2; Length 273;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
| | | | |  
DB 95 DGDGFA 100

RESULT 3  
E83946  
Pyruvate synthase beta subunit BH2373 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: E83946  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: E83946  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-288 <STO>  
A:Cross-references: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA06092.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2373

Query Match 100.0%; Score 34; DB 2; Length 288;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
| | | | |  
DB 92 DGDGFA 97

RESULT 4  
C95950  
conserved hypothetical protein SMB21202 [imported] - Sinorhizobium meliloti (strain 1021  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: C95950  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmaster, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: C95950  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-310 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC49267.1; PID:gl5140753; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMB21202  
A:Genome: plasmid  
C:Superfamily: Bacillus subtilis probable ABC transporter ydfo

Query Match 100.0%; Score 34; DB 2; Length 310;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
| | | | |

DB 123 DGDGFA 128

RESULT 5  
T35417  
probable beta-lactamase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35417  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: Z21577  
A:Accession: T35417  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-310 <OLI>  
A:Cross-references: EMBL:AL049485; PIDN:CAB39710.1; GSPDB:GN00070; SCOEDB:SC6A5.26C  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC6A5.26C

Query Match 100.0%; Score 34; DB 2; Length 310;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
| | | | |  
DB 295 DGDGFA 300

RESULT 6  
AC2878  
aromatic compounds dioxygenase [imported] - Agrobacterium tumefaciens (strain C58, Dupont  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AC2878  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AC2878  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-314 <KUR>  
A:Cross-references: GB:AF008688; PIDN:AAL43441.1; PID:gl7740944; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu2453  
A:Map position: circular chromosome  
C:Superfamily: Bacillus subtilis probable ABC transporter ydfo

Query Match 100.0%; Score 34; DB 2; Length 314;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
| | | | |  
DB 123 DGDGFA 128

RESULT 7  
E97654  
ABC transporter homolog ydfo [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C:Accession: E97654  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* strain 1594  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: E97654  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-325 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK88190.1; PID:gl5157638; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_4455  
A;Map position: circular chromosome  
C;Superfamily: *Bacillus subtilis* probable ABC transporter ydfO

Query Match 100.0%; Score 34; DB 2; Length 325;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
Db 134 DGDGFA 139

RESULT 8  
B91100  
hypothetical protein ECg3770 [imported] - *Escherichia coli* (strain O157:H7, substrain R1)  
C;Species: *Escherichia coli*  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: B91100  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: B91100  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-326 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA37193.1; PID:gl3363242; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain R1MD 0509952  
C;Genetics:  
A;Gene: ECS3770

Query Match 100.0%; Score 34; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
Db 78 DGDGFA 83

RESULT 9  
F85945  
hypothetical protein ygfZ [imported] - *Escherichia coli* (strain O157:H7, substrain EDL93)  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: F85945  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimallanta, E.; Potamouis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: F85945  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-326 <STO>  
A;Cross-references: GB:AE005174; NID:gl2517426; PIDN:AGS8026.1; GSPDB:GN00145; UMGF:242  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: ygfZ

Query Match 100.0%; Score 34; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
Db 78 DGDGFA 83

# RESULT 10 B65074

hypothetical protein b2898 - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: B65074

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B65074

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-326 <BLAT>

A;Cross-references: GB:AE000373; GB:U00096; NID:G2367173; PIDN:AAC75936.1; PID:gl789265;

A;Experimental source: strain K-12, substrain MGL655

Query Match 100.0%; Score 34; DB 2; Length 326;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
Db 78 DGDGFA 83

# RESULT 11 DB4325

Htr17 transducer [imported] - *Halobacterium* sp. NRC-1

C;Species: *Halobacterium* sp. NRC-1

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C;Accession: DB4325

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lie

A;Title: Genome sequence of *Halobacterium* species NRC-1.

A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: DB4325

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-536 <STO>

A;Cross-references: GB:AE004437; NID:gl0581193; PIDN:AAG19968.1; GSPDB:GN00138

C;Genetics:

A;Gene: htr17

C;Superfamily: *Halobacterium salinarum* transducer protein htrI

Query Match 100.0%; Score 34; DB 2; Length 536;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6  
|||||  
Db 358 DGDGFA 363

# RESULT 12 F84194

Htr14 transducer [imported] - *Halobacterium* sp. NRC-1

C;Species: *Halobacterium* sp. NRC-1

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C;Accession: F84194

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: F84194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <STO>  
A:Cross-references: GB:AE004437; NID:gl0579976; PIDN:AAG18922.1; GSPDB:GN00138  
C:Genetics:  
C:Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 100.0%; Score 34; DB 2; Length 627;  
Best Local Similarity 100.0%; Pred. No. 51; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6  
Db 399 DGDGFA 404

RESULT 13  
A84328  
Htr2 transducer [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A84328  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
; Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: A84328  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-764 <STO>  
A:Cross-references: GB:AE004437; NID:gl0581218; PIDN:AAG19989.1; GSPDB:GN00138  
C:Genetics:  
C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 100.0%; Score 34; DB 2; Length 764;  
Best Local Similarity 100.0%; Pred. No. 63; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6  
Db 602 DGDGFA 607

RESULT 14  
T44946  
transducer protein htrII [validated] - Halobacterium salinarum  
N:Alternate names: methyl-accepting taxis protein htrII; transducer protein htp2  
C:Species: Halobacterium salinarum  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 18-Aug-2000  
C:Accession: T44946  
R:Zhang, W.; Brooun, A.; Mueller, M.M.; Alam, M.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8230-8235, 1996  
A:Title: The primary structures in the Archaeon Halobacterium salinarum blue light rece  
A:Reference number: Z22877; MUID:96323203; PMID:8710852  
A:Accession: T44946  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-765 <ZHA>  
A:Cross-references: EMBL:U62676; PIDN:AAC44369.1  
A:Experimental source: strain Flx15  
A>Note: the source is designated as Halobacterium salinarum  
C:Genetics:  
A:Gene: htrII; htp2

C:Function:  
A:Description: involved in phototactic signal transduction from sensory rhodopsin II (PIF  
A>Note: the adaptation response is mediated by methylation of the transducer protein  
C:Superfamily: Halobacterium salinarum transducer protein htrII  
C:Keywords: methylated amino acid; signal transduction; transmembrane protein  
F:481-733/Region: MCP signalling domain similarity

Query Match 100.0%; Score 34; DB 1; Length 765;  
Best Local Similarity 100.0%; Pred. No. 63; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6  
Db 603 DGDGFA 608

RESULT 15  
T44597  
transducer protein htpVI [similarity] - Halobacterium salinarum  
N:Alternate names: methyl-accepting transducer protein htd; transducer protein htp6  
C:Species: Halobacterium salinarum  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T44597  
R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.  
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996  
A:Title: Signal transduction in the archaeon Halobacterium salinarum is processed throug  
A:Reference number: Z22804; MUID:96209786; PMID:8643458  
A:Accession: T44597  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-777 <ZHA>  
A:Cross-references: EMBL:U75438; NID:gl654424; PIDN:AAB17883.1; PID:gl654425  
A:Experimental source: strain mFlx15  
A>Note: the source is designated as Halobacterium salinarum  
C:Genetics:  
A:Gene: htpVI; htp6; htd  
C:Superfamily: Halobacterium salinarum transducer protein htrII  
C:Keywords: methylated amino acid; signal transduction; transmembrane protein  
F:494-746/Region: MCP signalling domain similarity

Query Match 100.0%; Score 34; DB 2; Length 777;  
Best Local Similarity 100.0%; Pred. No. 64; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6  
Db 616 DGDGFA 621

Search completed: June 1, 2004, 18:10:27  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 17:59:23 ; Search time 11 Seconds  
(without alignments)  
28.402 Million cell updates/sec

Title: US-09-715-763A-5  
Perfect score: 34  
Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	325	1 UP14_ECOLI	P39179 escherichia
2	34	100.0	763	1 HTR2_HALM1	Q9b81 halobacteri
3	34	100.0	764	1 HTR2_HALSA	P71410 halobacteri
4	34	100.0	778	1 HTR6_HALM1	Q9hr92 halobacteri
5	34	100.0	778	1 HTR6_HALSA	Q48319 halobacteri
6	31	91.2	535	1 HTR1_HALM1	P33741 halobacteri
7	31	91.2	535	1 HTR1_HALSA	P33955 halobacteri
8	30	88.2	78	1 POC7_PHLPR	O82040 phleum prat
9	30	88.2	80	1 POC7_CYNDA	P4092 cynodon dac
10	30	88.2	81	1 POC3_SYRVU	P38171 syringa vul
11	30	88.2	84	1 POC3_OLEBU	O81092 olea europ
12	30	88.2	85	1 POC4_ALNGL	O81701 alnus glut
13	30	88.2	85	1 POC4_BETVE	Q39419 betula verr
14	30	88.2	148	1 CALL_DROME	P49258 drosophila
15	30	88.2	148	1 CALM_METSE	P02596 metridium s
16	30	88.2	148	1 CALM_PATSP	P02595 patinopecte
17	30	88.2	214	1 HS30_ONCTS	P42931 oncorhynch
18	30	88.2	259	1 FKB7_HUMAN	Q9y680 homo sapien
19	30	88.2	331	1 RCN1_HUMAN	Q15293 homo sapien
20	30	88.2	340	1 YMD7_YEAST	Q03703 saccharomyc
21	30	88.2	389	1 TAL_GLOVI	Q7nk81 gloebacteri
22	30	88.2	391	1 TAL_SYNV3	P72797 synecocyst
23	30	88.2	402	1 PGL1_ERWCA	P18192 erwinia car
24	30	88.2	402	1 PGL2_ERWCA	P26509 erwinia car
25	30	88.2	433	1 HTR2_HALVA	P42258 haloarcula
26	30	88.2	438	1 PORO_PGEAE	P32977 pseudomonas
27	30	88.2	444	1 GID_STR8E	Q8dgs1 streptococc
28	30	88.2	468	1 AMYE_BACFI	P96513 bacillus fi
29	30	88.2	472	1 RE16_SCHPO	P40379 schizosacch
30	30	88.2	477	1 FES_FSVST	P00543 feline sarc
31	30	88.2	494	1 TCWN_STRGA	P16559 streptomyc
32	30	88.2	529	1 HSF1_HUMAN	Q00613 homo sapien
33	30	88.2	559	1 PMG1_MAIZE	P30792 zea mays (m

## ALIGNMENTS

## RESULT 1

UP14\_ECOLI  
ID UP14\_ECOLI STANDARD; PRT; 325 AA.  
AC P39179; Q45826;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Unknown protein from 2D-page (Spot PR51).  
GN YGFZ OR B2898 OR SF2884 OR S3083.  
OS Escherichia coli, and  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
RX MEDLINE=974426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE OF 1-12.  
RC SPECIES=E.coli; STRAIN=K12 / BMG2;  
RX MEDLINE=97443975; PubMed=9298646;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
in the genome of Escherichia coli K-12.";  
Electrophoresis 18:1259-1313(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=2272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
through comparison with genomes of Escherichia coli K12 and O157.";  
Nucleic Acids Res. 30:4432-4441(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T.";  
Infect. Immun. 71:2775-2786(2003).  
RN [5]

34 30 88.2 609 1 FES\_PSVGA P00542 feline sarc  
35 30 88.2 640 1 GATE\_METKA Q8tz12 methanopyru  
36 30 88.2 756 1 EFER\_HUMAN O75154 homo sapien  
37 30 88.2 774 1 GLBI\_STRCO Q59833 streptomyc  
38 30 88.2 820 1 FES\_FELCA P14238 felis silve  
39 30 88.2 1196 1 AMYB\_PAEPO P21543 paenibacill  
40 30 88.2 1901 1 YZ08\_WYCTU O53553 mycobacteri  
41 29 85.3 146 1 KLA2\_ECOLI P52603 escherichia  
42 29 85.3 545 1 HTRS\_HALM1 Q48318 halobacteri  
43 28 82.4 154 1 PTGA\_MYCCA P45618 mycoplasma  
44 28 82.4 189 1 STA\_STRLA P08457 streptomyc  
45 28 82.4 189 1 YZ05\_ADE07 P05665 human adeno

Wed Jun 2 16:48:17 2004

us-09-715-763a-5.rsp

```

IDENTIFICATION BY MASS SPECTROMETRY.
RP SPECIES=E.coli;
RC MEDLINE=99420866; PubMed=10493123;
RX Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RA "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -!- SIMILARITY: STRONG, TO B.APHIDICOLA (SUBSP. ACYRTHOSIPHON PISUM)
CC BUA35. ALSO TO H.INFLUENZAE H10466.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U28375; AA83079.1; -.
CC PIR; A84328; A84328.
CC HSSP; P02942; 1Q07.
CC InterPro; IPR004089; Chmtaxis_transd.
CC Pfam; PF00672; HAMP; 1.
CC Pfam; PF00015; MCPsignal; 1.
CC SMART; SM00304; HAMP; 2.
CC SMART; SM00283; MA; 1.
CC PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
CC PROSITE; PS00885; HAMP; 2.
KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation;
KW Complete proteome; Repeat.
KW INIT_MET 0 0
FT DOMAIN 1 15
FT TRANSMEM 16 36
FT DOMAIN 37 277
FT TRANSMEM 278 297
FT DOMAIN 298 763
FT TRANSMEM 302 354
FT DOMAIN 356 449
FT TRANSMEM 468 704
FT DOMAIN 763 AA; CF7A8FF04DFF309A CRC64;
SQ SEQUENCE 325 AA; 35963 MW; 333889EB37048A36 CRC64;
Query Match 100.0%; Score 34; DB 1; Length 325;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGDGFA 6
Db 77 DGDGFA 82
-----
RESULT 2
HTR2 HALN1
ID HTR2 HALN1 STANDARD; PRT; 763 AA.
AC Q9HP81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II).
DE HTR2 OR VNG1765G.
GN Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitthausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Bhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Transduces signals from the phototaxis receptor sensory
CC rhodopsin II (SR-II) to the flagellar motor. Responds to light
CC changes through the variation of the level of methylation. Also
CC acts as a chemotransducer (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
-----
-!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
-!- SIMILARITY: Contains 2 HAMP domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE005080; AAG19989.1; -.
CC PIR; A84328; A84328.
CC HSSP; P02942; 1Q07.
CC InterPro; IPR004089; Chmtaxis_transd.
CC Pfam; PF00672; HAMP; 1.
CC Pfam; PF00015; MCPsignal; 1.
CC SMART; SM00304; HAMP; 2.
CC SMART; SM00283; MA; 1.
CC PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
CC PROSITE; PS00885; HAMP; 2.
KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation;
KW Complete proteome; Repeat.
KW INIT_MET 0 0
FT DOMAIN 1 15
FT TRANSMEM 16 36
FT DOMAIN 37 277
FT TRANSMEM 278 297
FT DOMAIN 298 763
FT TRANSMEM 302 354
FT DOMAIN 356 449
FT TRANSMEM 468 704
FT DOMAIN 763 AA; CF7A8FF04DFF309A CRC64;
SQ SEQUENCE 763 AA; 78911 MW;
Query Match 100.0%; Score 34; DB 1; Length 763;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGDGFA 6
Db 601 DGDGFA 606
-----
RESULT 3
HTR2 HALSA
ID HTR2 HALSA STANDARD; PRT; 764 AA.
AC F71410;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II).
DE HTR2 OR HTRII.
GN Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FLX15;
RX MEDLINE=96323203; PubMed=8710852;
RA Zhang W., Brooun A., Mueller M.M., Alam M.;
RA "The primary structures of the Archaeon Halobacterium salinarum blue
RT light receptor sensory rhodopsin II and its transducer, a methyl-
RT accepting protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).
RN [2]
RP FUNCTION.
RX PubMed=9515936;
RA Hou S., Brooun A., Yu H.S., Freitas T., Alam M.;
RA "Sensory rhodopsin II transducer HtrII is also responsible for serine
RT chemotaxis in the archaeon Halobacterium salinarum.";

```

J. Bacteriol. 180:1600-1602(1998).  
 -!- FUNCTION: Transduces signals from the phototaxis receptor sensory rhodopsin II (SR-II) to the flagellar motor. Responds to light changes through the variation of the level of methylation. Also acts as a chemotransducer.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.  
 -!- SIMILARITY: Contains 2 HAMP domains.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; U62676; AAC44369.1; --  
 DR PIR; T44946; T44946.  
 DR HSSP; P02942; 1Q07.  
 DR InterPro; IPR004089; Chmtaxis\_transd.  
 DR InterPro; IPR003660; HAMP.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR SMART; SM00283; NA; 1.  
 DR PROSITE; PS01111; CHEMOTAXIS\_TRANSDUC\_2; 1.  
 DR PROSITE; PS50885; HAMP; 2.  
 KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation;  
 Repeat.  
 FT INIT MET 0 0 BY SIMILARITY.  
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 278 298 POTENTIAL.  
 FT DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 303 355 HAMP 1.  
 FT DOMAIN 397 450 HAMP 2.  
 FT DOMAIN 469 705 METHYL-ACCEPTING TRANSDUCER.  
 SQ SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460FC588 CRC64;  
 Query Match 100.0%; Score 34; DB 1; Length 764;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DGDGFA 6  
 Db 602 DGDGFA 607  
 |||||  
 RESULT 4  
 HTR6 HALN1  
 ID HTR6 HALN1 STANDARD; PRT; 778 AA.  
 AC Q9HR32;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Halobacterial transducer protein VI.  
 GN HTR4 OR HTPVI OR VNG0906G.  
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M., Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J., Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A., Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W., Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H., Isenbarger T.A., Peck R.P., Pohlschroder M., Spudis J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

EBhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 "Genome sequence of Halobacterium species NRC-1.";  
 Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 -!- FUNCTION: Potentially involved in chemo- or phototactic signal transduction (By similarity).  
 -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.  
 -!- SIMILARITY: Contains 2 HAMP domains.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AE005022; AAG19266.1; --  
 DR PIR; F84237; F84237.  
 DR HSSP; P02942; 1Q07.  
 DR InterPro; IPR004089; Chmtaxis\_transd.  
 DR InterPro; IPR003660; HAMP.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR PRINTS; PR00260; CHEMTRNSDUCR.  
 DR SMART; SM00304; HAMP; 2.  
 DR SMART; SM00283; NA; 1.  
 DR PROSITE; PS01111; CHEMOTAXIS\_TRANSDUC\_2; 1.  
 DR PROSITE; PS50885; HAMP; 2.  
 KW Transducer; Transmembrane; Complete proteome; Repeat.  
 FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 27 47 POTENTIAL.  
 FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 297 317 POTENTIAL.  
 FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 318 370 HAMP 1.  
 FT DOMAIN 412 465 HAMP 2.  
 FT DOMAIN 484 720 METHYL-ACCEPTING TRANSDUCER.  
 SQ SEQUENCE 778 AA; 82076 MW; D752278727A4FA79 CRC64;  
 Query Match 100.0%; Score 34; DB 1; Length 778;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DGDGFA 6  
 Db 617 DGDGFA 622  
 |||||  
 RESULT 5  
 HTR6 HALSA  
 ID HTR6 HALSA STANDARD; PRT; 778 AA.  
 AC Q48319;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Halobacterial transducer protein VI.  
 GN HTR4 OR HTPVI.  
 OS Halobacterium salinarum.  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=2242;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S9;  
 RX MEDLINE=96275896; PubMed=8674984;  
 RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K., Oesterheld D.;  
 "A family of halobacterial transducer proteins.";  
 FEBS Microbiol. Lett. 139:161-168(1996).  
 -!- FUNCTION: Potentially involved in chemo- or phototactic signal transduction.  
 -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.

Wed Jun 2 16:48:17 2004

us-09-715-763a-5.rsp

```

CC CC -1- SIMILARITY: Contains 2 HAMP domains.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X95590; CRA64843.1; -.
CC CC PIR; T48897; T48897.
CC CC HSSP; P02942; LQ07.
CC CC InterPro; IPR004089; Chmtaxis_trans.
CC CC InterPro; IPR003660; HAMP.
CC CC InterPro; IPR004090; Me_chemotaxis.
CC CC Pfam; PF00672; HAMP; 1.
CC CC Pfam; PF00015; MCPsignal; 1.
CC CC PRINTS; PR00260; CHEMTRNSDUCR.
CC CC SMART; SM00304; HAMP; 2.
CC CC SMART; SM00283; MA; 1.
CC CC PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
CC CC Transducer; Transmembrane; Repeat.
CC CC DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
CC CC TRANSMEM 27 47 POTENTIAL.
CC CC DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).
CC CC TRANSMEM 297 317 POTENTIAL.
CC CC DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).
CC CC DOMAIN 318 370 HAMP 1.
CC CC DOMAIN 412 465 HAMP 2.
CC CC DOMAIN 484 720 METHYL-ACCEPTING TRANSDUCER.
CC CC SEQUENCE 778 AA; 82077 MW; 134C7D7F0A33334CD CRC64;

Query Match 100.0%; Score 34; DB 1; Length 778;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
DB 617 DGDGFA 622

RESULT 6
HTR1_HALN1
ID HTR1_HALN1 STANDARD; PRT; 535 AA.
AC P33741; Q9HPP6;
AT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE protein I) (MPP-I).
GN HTR1 OR HTR1 OR VNG1659G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and
OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
ON NCBI_TaxID=64091, 2242;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger I.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.
RC SPECIES=H.halobium; STRAIN=FLX5R;
RX MEDLINE=93101637; PubMed=1465418;
RA Yao V.J., Spudich J.L.;
RA "Primary structure of an archaeobacterial transducer, a
RT methyl-accepting protein associated with sensory rhodopsin I.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).
CC CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC CC SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC CC -1- SIMILARITY: Contains 2 HAMP domains.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB005075; AAG19913.1; -.
CC CC PIR; A47190; A47190.
CC CC PIR; E84318; E84318.
CC CC HSSP; P02942; LQ07.
CC CC InterPro; IPR004089; Chmtaxis_trans.
CC CC InterPro; IPR003660; HAMP.
CC CC InterPro; IPR004090; Me_chemotaxis.
CC CC Pfam; PF00672; HAMP; 1.
CC CC Pfam; PF00015; MCPsignal; 1.
CC CC PRINTS; PR00260; CHEMTRNSDUCR.
CC CC SMART; SM00304; HAMP; 2.
CC CC SMART; SM00283; MA; 1.
CC CC PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
CC CC Transducer; Photoreceptor; Transmembrane; Methylation;
CC CC Complete proteome; Repeat.
CC CC INIT MET 0 0 CYTOPLASMIC (POTENTIAL).
CC CC DOMAIN 1 13 POTENTIAL.
CC CC TRANSMEM 14 28 EXTRACELLULAR (POTENTIAL).
CC CC DOMAIN 29 38 POTENTIAL.
CC CC TRANSMEM 39 54 CYTOPLASMIC (POTENTIAL).
CC CC DOMAIN 55 535 HAMP 1.
CC CC DOMAIN 54 106 HAMP 2.
CC CC DOMAIN 148 201 METHYL-ACCEPTING TRANSDUCER.
CC CC DOMAIN 220 458 METHYLATION.
CC CC MOD RES 265 265 METHYLATION.
CC CC MOD RES 272 272 METHYLATION.
CC CC MOD RES 279 279 METHYLATION.
CC CC MOD RES 463 463 METHYLATION.
CC CC MOD RES 472 472 METHYLATION.
CC CC SEQUENCE 535 AA; 56544 MW; B9945E4F66A9D091 CRC64;

Query Match 91.2%; Score 31; DB 1; Length 535;
Best Local Similarity 83.3%; Pred. No. 11e-02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
DB 355 DGDGFS 360

RESULT 7
HTR1_HALSA
ID HTR1_HALSA STANDARD; PRT; 535 AA.
AC P33955;
AT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE protein I) (MPP-I).
GN HTR1 OR HTR1 OR HTR.

```

```

OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
ON NCBI_TaxID=2242;
RX STRAIN=L33;
RX MEDLINE=90384955; PubMed=2205842;
RA Ferrando E., Krah M., Marwan W., Oesterhelt D.;
RT "A gene from S. pombe with homology to E. coli RNase III blocks
RL conjugation and sporulation when overexpressed in wild type cells.";
RL Nucleic Acids Res. 18:5304-5304(1990).
RN [2]
RP FUNCTION.
RP MEDLINE=94244615; PubMed=8187768;
RX Krah M., Marwan W., Vermeglio A., Oesterhelt D.;
RA "Phototaxis of Halobacterium salinarum requires a signalling complex
RT of sensory rhodopsin I and its methyl-accepting transducer HtrI.";
RL EMBO J. 13:2150-2155(1994).
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68591; CAA48578.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me.chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
KW Transducer; Photoreceptor; Transmembrane; Methylation; Repeat.
FT INIT MET 0 0
FT DOMAIN 1 8
FT TRANSMEM 9 29
FT DOMAIN 30 33
FT TRANSMEM 34 54
FT DOMAIN 55 535
FT DOMAIN 148 201
FT DOMAIN 220 458
FT MOD_RES 265 265
FT MOD_RES 272 272
FT MOD_RES 279 279
FT MOD_RES 463 463
FT MOD_RES 472 472
SQ SEQUENCE 535 AA; 56814 MW; 504A165A47FA8A45 CRC64;

Query Match 91.2%; Score 31; DB 1; Length 535;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 355 DGDGFS 360
|||||
TISSUE=Pollen;

RESULT 8
POC7_PHLPR
ID POC7_PHLPR STANDARD; PRT; 78 AA.
AC O82040;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Polcalcic Phl p 7 (Calcium-binding pollen allergen Phl p 7) (P7).
OS Phleum pratense (Common timothy).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Aveneae; Phleum.
ON NCBI_TaxID=15957;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Pollen;
RX MEDLINE=99242550; PubMed=10224228;
RA Niederberger V., Hayek B., Vrtala S., Laffer S., Twardosz A.,
RA Vangelista L., Sperr W.R., Valent P., Rumpold H., Kraft D.,
RA Ehrenberger K., Valenta R., Spitzauer S.;
RT "Calcium-dependent immunoglobulin E recognition of the apo- and
RT calcium-bound form of a cross-reactive two EF-hand timothy grass
RT pollen allergen, Phl p 7.";
RL FASEB J. 13:843-856(1999).
CC -!- ALLERGEN: Causes an allergic reaction in human. Binds IGE.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y17835; CAA76887.1; -.
DR HSSP; P25070; 1AVJ.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001125; Recoverin.
DR Pfam; PF00036; efhand; 2.
DR PRINTS; PR00450; RECOVERIN.
DR PRODOM; PD000012; EF-hand; 1.
DR PROSITE; PS00054; EFh; 2.
DR PROSITE; PS00018; EF HAND; 2.
KW Calcium-binding; Repeat; Allergen.
FT CA_BIND 13 24 EF-HAND 1 (POTENTIAL).
FT CA_BIND 48 59 EF-HAND 2 (POTENTIAL).
FT SEQUENCE 78 AA; 8677 MW; 1469370AAEAE2244 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5
Db 50 DGDGF 54
|||||
TISSUE=Pollen;

RESULT 9
POC7_CYNDA
ID POC7_CYNDA STANDARD; PRT; 80 AA.
AC P94032; Q9SAR4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Polcalcic Cyn d 7 (Calcium-binding pollen allergen Cyn d 7) (Calcium
DE binding protein B1).
OS Cynodon dactylon (Bermuda grass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Chloridoideae; Cynodonteae; Cynodon.
ON NCBI_TaxID=28909;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Pollen;

```

Wed Jun 2 16:48:17 2004

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF078681; AAK01144.1; -.
CC HSSP; P02588; 1AIVS.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF000036; ehand; 2.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00054; EFh; 2.
CC PROSITE; PS00018; EF_HAND; 2.
CC Calcium-binding; Repeat; Allergen.
CC EF-HAND 1 (POTENTIAL).
CC CA_BIND 16 27
CC CA_BIND 51 62
CC SEQUENCE 81 AA; 8994 MW; 9C0C4A8BCE03B3FA CRC64;
CC -----
CC Query Match 88.2%; Score 30; DB 1; Length 81;
CC Best Local Similarity 100.0%; Pred. No. 26;
CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 1 DGDGF 5
DB 53 DGDGF 57

RESULT 11
POC3_SYRVU STANDARD; PRT; 84 AA.
ID POC3_SYRVU STANDARD; PRT; 84 AA.
AC O81092;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polcalcine Ole e 3 (Calcium-binding pollen allergen Ole e 3).
GN OLE3.
OS Olea europaea (Common olive).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Oleaceae; Olea.
OX NCBI_TaxID=4146;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99089607; PubMed=9874211;
RA Ledesma A., Villalba M., Batanero E., Rodriguez R.;
RT "Molecular cloning and expression of active Ole e 3, a major allergen
RT from olive-tree pollen and member of a novel family of Ca2+-binding
RT proteins (polcalcins) involved in allergy.";
RL Eur. J. Biochem. 258:454-459(1998).
CC -|- ALLERGEN: Causes an allergic reaction in human. Binds IgE.
CC -|- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91256; CAA62634.1; ALT_INIT.
CC EMBL; U35683; AAC49648.1; ALT_INIT.
CC EMBL; U75585; AAD00247.1; -.
CC HSSP; P25070; 1AVJ.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 2.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00054; EFh; 2.
CC PROSITE; PS00018; EF_HAND; 2.
CC Calcium-binding; Repeat; Allergen.
CC EF-HAND 1 (POTENTIAL).
CC CA_BIND 15 26
CC CA_BIND 50 61
CC SEQUENCE 80 AA; 8801 MW; F83347BD7C2676F5 CRC64;
CC -----
CC Query Match 88.2%; Score 30; DB 1; Length 80;
CC Best Local Similarity 100.0%; Pred. No. 26;
CC Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 1 DGDGF 5
DB 52 DGDGF 56

RESULT 10
POC3_SYRVU STANDARD; PRT; 81 AA.
AC P58171;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polcalcine Syr v 3 (Calcium-binding pollen allergen Syr v 3).
GN SYRV3.
OS Syringa vulgaris (Common lilac).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Lamiales; Oleaceae; Syringa.
OX NCBI_TaxID=34270;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Pollen;
RA Ledesma A., Villalba M., Rodriguez R.;
RT "Cloning and expression of a protein homologous to Ole e 3 from
RT Syringa vulgaris.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -|- ALLERGEN: Causes an allergic reaction in human.
CC -|- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

Query Match 88.2%; Score 30; DB 1; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
 |||||  
 Db 56 DGDGF 60

RESULT 12  
 POC4 ALNGL STANDARD; PRT; 85 AA.  
 AC O81701;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polcalcin Aln g 4 (Calcium-binding pollen allergen Aln g 4).  
 OS Alnus glutinosa (Alder).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fagales; Betulaceae; Alnus.  
 OX NCBI\_TaxID=3517;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Pollen;  
 RX MEDLINE=99077576; PubMed=9862740;  
 RA Hayek B., Vangelista L., Pastore A., Sperr W.R., Valent P., Vrtala S.,  
 RA Niederberger V., Twardosz A., Kraft D., Valenta R.;  
 RT "Molecular and immunologic characterization of a highly cross-reactive  
 RT two EF-hand calcium-binding alder pollen allergen, Aln g 4;  
 RT structural basis for calcium-modulated IgE recognition.";  
 RL J. Immunol. 161:7031-7039 (1998).  
 CC -!- MASS SPECTROMETRY; MW=9318.6; METHOD=MALDI.  
 CC -!- ALLERGEN: Causes an allergic reaction in human.  
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; Y17713; CAA76831.1; --  
 HSP; P02593; 1AK8.  
 InterPro; IPR002048; EF-hand.  
 Pfam; PF00036; eFhand; 2.  
 ProDom; PD000012; EF-hand; 1.  
 SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF-hand; 2.  
 DR CA BIND 20 31 EF-HAND 1 (POTENTIAL).  
 FT CA BIND 55 66 EF-HAND 2 (POTENTIAL).  
 SQ SEQUENCE 85 AA; 9362 MW; 903774CCIF714FAD CRC64;

Query Match 88.2%; Score 30; DB 1; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
 |||||  
 Db 57 DGDGF 61

RESULT 13  
 POC4 BETVE STANDARD; PRT; 85 AA.  
 AC Q39419; O04131;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Polcalcin Bet v 4 (Calcium-binding pollen allergen Bet v 4).  
 GN BETV4.  
 OS Betula verrucosa (White birch) (Betula pendula).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fagales; Betulaceae; Betula.  
 OX NCBI\_TaxID=3505;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pollen;  
 RX MEDLINE=98019240; PubMed=9353329;  
 RA Engel E., Richter K., Obermeyer G., Briza P., Kungl A.J., Simon B.,  
 RA Auer M., Ebner C., Rheinberger H.J., Breitenbach M., Ferreira F.;  
 RT "Immunological and biological properties of Bet v 4, a novel birch  
 RT pollen allergen with two EF-hand calcium-binding domains.";  
 RL J. Biol. Chem. 272:28630-28637 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pollen;  
 RX MEDLINE=98005106; PubMed=9345295;  
 RA Twardosz A., Hayek B., Seiberler S., Pastore A., Vangelista L.,  
 RA Groenlund H., Kraft D., Valenta R.;  
 RT "Molecular characterization, expression in *Escherichia coli*, and  
 RT epitope analysis of a two EF-hand calcium-binding birch pollen  
 RT allergen, Bet v 4.";  
 RL Biochem. Biophys. Res. Commun. 239:197-204 (1997).  
 CC -!- ALLERGEN: Causes an allergic reaction in human.  
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X87153; CAA60628.1; --  
 EMBL; Y12560; CAA73147.1; --  
 PIR; JCS711; JCS711.  
 PIR; S54819; S54819.  
 HSP; P02593; 1CDM.  
 InterPro; IPR002048; EF-hand.  
 Pfam; PF00036; eFhand; 2.  
 ProDom; PD000012; EF-hand; 1.  
 SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF-hand; 2.  
 DR CA BIND 20 31 EF-HAND 1 (POTENTIAL).  
 FT CA BIND 55 66 EF-HAND 2 (POTENTIAL).  
 FT CONFLICT 71 71 G->A (IN REF. 2).  
 SQ SEQUENCE 85 AA; 9447 MW; E2FC862E5B8ADFF CRC64;

Query Match 88.2%; Score 30; DB 1; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGF 5  
 |||||  
 Db 57 DGDGF 61

RESULT 14  
 CALL DROME STANDARD; PRT; 148 AA.  
 ID P49258; Q9VBL9;  
 AC P49258; Q9VBL9;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Calmodulin related protein 97A (Androcam protein).  
 GN AND OR CAMR97A OR CGI7769.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;



Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 [1]  
 NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95071221; PubMed=7980384;  
 RA Fyberg C., Parker H., Hutchison B., Fyberg E.A.;  
 RA "Drosophila melanogaster genes encoding three troponin-C isoforms and  
 RA a calmodulin-related protein.";  
 RL Biochem. Genet. 32:119-135(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bhandari D., Bolshakov S.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Brothier P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X76045; CAA53630.1; -;  
 CC EMBL; AE003753; AAF56511.1; -;  
 CC HSP; P02593; 1FW4.  
 CC FlyBase; FBgn0011273; And.  
 CC GO; GO:0005737; C:cytoplasm; IDA.  
 CC GO; GO:0005634; C:nucleus; IDA.  
 CC InterPro; IPR002048; EF-hand.  
 CC Pfam; PF00036; ehand; 4.  
 CC ProDom; PD000012; EF-hand; 2.  
 CC SMART; SM00054; EFh; 4.

DR PROSITE; PS00018; EF\_HAND; 2.  
 KW Calcium-binding; Repeat.  
 FT CA\_BIND 20 31 EF-HAND 1.  
 FT CA\_BIND 56 67 EF-HAND 2 (POTENTIAL).  
 FT CA\_BIND 93 104 EF-HAND 3.  
 FT CA\_BIND 129 140 EF-HAND 4.  
 FT CA\_BIND 105 105 L -> I (IN REF. 1).  
 SQ CONFLICT 148 AA; 17015 MW; 32DB8DC4B35C6CB8 CRC64;  
 FT SEQUENCE 148 AA; 17015 MW; 32DB8DC4B35C6CB8 CRC64;  
 Query Match 88.2%; Score 30; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGF 5  
 Db 95 DGDGF 99  
 RESULT 15  
 CALM METSE  
 ID CALM METSE STANDARD; PRT; 148 AA.  
 AC P02596;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Calmodulin.  
 OS Metridium senile (Brown sea anemone) (Frilled sea anemone), and  
 OS Renilla reniformis (Sea pansy).  
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;  
 OC Nynanthae; Metridiidae; Metridium.  
 OC NCBI\_TaxID=6116, 6136;  
 RN [1]  
 RN SEQUENCE.  
 RP MEDLINE=81062446; PubMed=6108109;  
 RX Takagi T., Nemoto T., Konishi K., Yazawa M., Yagi K.;  
 RA "The amino acid sequence of the calmodulin obtained from sea anemone  
 RT (metridium senile) muscle.";  
 RL Biochem. Biophys. Res. Commun. 96:377-381(1980).  
 RN [2]  
 RN SEQUENCE.  
 RP MEDLINE=81205529; PubMed=6263143;  
 RX Jamieson G.A. Jr., Bronson D.D., Schachar F.H., Vanaman T.C.;  
 RA "Structure and function relationships among calmodulins and troponin  
 RT C-like proteins from divergent eukaryotic organisms.";  
 RL Ann. N.Y. Acad. Sci. 356:11-13(1980).  
 CC -!- FUNCTION: Calmodulin mediates the control of a large number of  
 CC enzymes by Ca(++) . Among the enzymes to be stimulated by the  
 CC calmodulin-Ca(++) complex are a number of protein kinases and  
 CC phosphatases.  
 CC -!- MISCELLANEOUS: This protein has four functional calcium-binding  
 CC sites.  
 CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 DR PIR; A90036; MCJZR.  
 DR PIR; A90223; MCXAM.  
 DR HSP; P02593; 1AK8.  
 DR InterPro; IPR02048; EF-hand.  
 DR Pfam; PF00036; ehand; 4.  
 DR ProDom; PD000012; EF-hand; 2.  
 DR SMART; SM00054; EFh; 4.  
 DR PROSITE; PS00018; EF\_HAND; 4.  
 KW Calcium-binding; Repeat; Methylation.  
 FT MOD\_RES 1 115 115 METHYLATION (TRI-).  
 FT MOD\_RES 115 115 115 METHYLATION (TRI-).  
 FT CA\_BIND 20 31 EF-HAND 1.  
 FT CA\_BIND 56 67 EF-HAND 2.  
 FT CA\_BIND 93 104 EF-HAND 3.  
 FT CA\_BIND 129 140 EF-HAND 4.  
 SQ SEQUENCE 148 AA; 16708 MW; 4CEEBE8C4D750AA CRC64;  
 Query Match 88.2%; Score 30; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 1 DGDGF 5  
| | | | |  
Db 95 DGDGF 99

Search completed: June 1, 2004, 18:09:01  
Job time : 12 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 18:05:23 ; Search time 39 Seconds  
(without alignments)  
48.541 Million cell updates/sec

Title: US-09-715-763A-5  
Perfect score: 34  
Sequence: 1 DGDGFA 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_protist.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	230	Q8EY42	Q8EY42 leptospira
2	34	100.0	273	Q25312	Q25312 helicobacte
3	34	100.0	273	Q9ZLP0	Q9ZLP0 helicobacte
4	34	100.0	274	Q68229	Q68229 helicobacte
5	34	100.0	276	Q9P9A7	Q9P9A7 uncultured
6	34	100.0	288	Q9KAB6	Q9KAB6 bacillus ba
7	34	100.0	288	Q81WQ9	Q81WQ9 bacillus an
8	34	100.0	288	Q81A22	Q81A22 bacillus ce
9	34	100.0	310	Q98FP0	Q98FP0 rhizobium l
10	34	100.0	310	Q92V50	Q92V50 rhizobium m
11	34	100.0	310	Q9X7X2	Q9X7X2 streptomyc
12	34	100.0	325	Q8UCN0	Q8UCN0 agrobacteri
13	34	100.0	326	Q8XD41	Q8XD41 escherichia
14	34	100.0	326	Q8FE70	Q8FE70 escherichia
15	34	100.0	536	Q9HPA2	Q9HPA2 halobacteri
16	34	100.0	539	Q7XX03	Q7XX03 oryza sativ

17	34	100.0	627	17	Q9HS86
18	34	100.0	777	1	P71415
19	34	100.0	788	1	O93643
20	34	100.0	789	17	Q9HRA1
21	34	100.0	948	16	Q8XX42
22	34	100.0	985	16	Q8XUP9
23	34	100.0	1351	16	Q8XZ58
24	34	100.0	3770	16	Q87G75
25	31	91.2	133	7	Q860H6
26	31	91.2	195	10	Q84Y42
27	31	91.2	233	2	Q9L3F3
28	31	91.2	262	12	Q89340
29	31	91.2	267	16	Q9HZ01
30	31	91.2	279	2	Q8RJ09
31	31	91.2	288	16	Q99U13
32	31	91.2	288	16	Q8CPF2
33	31	91.2	347	17	Q97BL3
34	31	91.2	353	16	Q89V56
35	31	91.2	372	7	Q31406
36	31	91.2	374	5	Q9N6B1
37	31	91.2	433	16	Q92TA1
38	31	91.2	475	16	Q82CQ6
39	31	91.2	481	16	Q82EY7
40	31	91.2	486	5	Q8SYF5
41	31	91.2	489	16	Q82N21
42	31	91.2	532	16	Q9L1J5
43	31	91.2	560	16	Q8FAI5
44	31	91.2	576	16	Q9L1I5
45	31	91.2	635	2	Q84HK0

#### ALIGNMENTS

#### RESULT 1

Q8EY42 PRELIMINARY; PRT; 230 AA.  
AC Q8EY42;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Two-component response regulator.  
GN LB015.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RW [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar Lai;  
RA Ren S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE011590; AAN51574.1; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0000156; F:two-component response regulator activity; IEA.  
DR GO; GO:0000160; P:two-component signal transduction system (p. . .); IEA.  
DR InterPro; IPR001789; Response reg.  
DR InterPro; IPR001867; Trans\_reg\_C.  
DR Pfam; PF00072; response\_reg\_C.  
DR Pfam; PF00486; trans\_reg\_C; 1.  
DR ProDom; PD000039; Response reg; 1.  
DR ProDom; PD000329; Trans\_reg\_C; 1.  
DR SMART; SMO0448; REC\_1; RESPONSE\_REGULATORY; 1.  
DR PROSITE; PS0110; RESPONSE\_REGULATORY; 1.  
DR Complete proteome.  
SQ SEQUENCE 230 AA; 26523 MW; 0C594D9D2E28E6AE CRC64;

Query Match 100.0%; Score 34; DB 16; Length 230;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGDGFA 6

Db 58 DGDGFA 63

```

SQ SEQUENCE 273 AA; 30429 MW; A98A9BE5D4DB634E CRC64;

Query Match 100.0%; Score 34; DB 16; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 95 DGDGFA 100
|||||

RESULT 4
O68229 PRELIMINARY; PRT; 274 AA.
ID O68229;
AC O68229;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE OorB subunit of 2-oxoglutarate:acceptor oxidoreductase.
GN COOR.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 11637;
RX MEDLINE=95332198; PubMed=7608066;
RA Hughes N.J., Chalk P.A., Clayton C.L., Kelly D.J.;
RT "Identification of carboxylation enzymes and characterization of a
RT novel four-subunit pyruvate:flavodoxin oxidoreductase from
RT Helicobacter pylori.";
RL J. Bacteriol. 177:3953-3959 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 11637;
RX MEDLINE=98155137; PubMed=9495749;
RA Hughes N.J., Clayton C.L., Chalk P.A., Kelly D.J.;
RT "Helicobacter pylori porCAB and oordABC genes encode distinct
RT pyruvate:flavodoxin and 2-oxoglutarate:acceptor oxidoreductases which
RT mediate electron transport to NADP.";
RL J. Bacteriol. 180:1119-1128 (1998).
DR EMBL; AF021094; AAC38212.1; -.
SQ SEQUENCE 274 AA; 30645 MW; BD161B857881AB97 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 95 DGDGFA 100
|||||

RESULT 5
O9P9A7 PRELIMINARY; PRT; 276 AA.
ID O9P9A7;
AC O9P9A7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Metal-dependent hydrolase.
OS uncultured marine group II euryarchaeote 37Fl1.
OC Archaea; Euryarchaeota; Marine Group II; environmental samples.
OX NCBI_TaxID=133822;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21128653; PubMed=11233160;
RA Beja O., Suzuki M.T., Koonin E.V., Aravind L., Hadd A., Nguyen L.P.,
RA Villacorta R., Amjadi M., Garrigues C., Jovanovich S.B., Feldman R.A.,
RA Delong E.F.;
RT "Construction and analysis of bacterial artificial chromosome
RT libraries from a marine microbial assemblage.";

```

RL Environ. Microbiol. 2:516-529 (2000).  
 DR EMBL: AF268611; AAF97210.1; -.  
 DR GO: 0016787; F:hydroxylase activity; IEA.  
 DR InterPro: IPR001279; Blackmase-like.  
 DR Pfam: PF00753; lactamase\_B; 1.  
 KW Hydroxylase.  
 SQ SEQUENCE 276 AA; 30145 MW; ADECD6DBE4F6A74F CRC64;

Query Match 100.0%; Score 34; DB 1; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 259 DGDGFA 264

# RESULT 6

ID O9KAB6 PRELIMINARY; PRT; 288 AA.

AC O9KAB6;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Pyruvate synthase beta subunit.

GN BH2373.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331 (2000).

DR EMBL: AP001515; BAB06092.1; -.

DR PIR: E83946; E83946.

KW Complete proteome.

SQ SEQUENCE 288 AA; 31291 MW; 3C29FBBFF757CCE8 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 288;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 92 DGDGFA 97

# RESULT 7

ID O81WQ9 PRELIMINARY; PRT; 288 AA.

AC O81WQ9;

DT 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Pyruvate ferredoxin oxidoreductase, beta subunit, putative.

GN BA3909.

OS Bacillus anthracis (strain Ames).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=198094;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608414; PubMed=12721629;

RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,

RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,

RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,

RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,

RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,

RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
 RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolisto A.-B.,  
 RA Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86 (2003).  
 DR EMBL: AE017036; AAP27642.1; -.  
 DR TIGR: BA3909; -.  
 KW Pyruvate; Complete proteome.  
 SQ SEQUENCE 288 AA; 31365 MW; 9C5FAED35BF46039 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 288;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 92 DGDGFA 97

# RESULT 8

ID O81A22

AC O81A22 PRELIMINARY; PRT; 288 AA.

DT 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Pyruvate synthase subunit porB (EC 1.2.7.1).

GN BC3773.

OS Bacillus cereus (strain ATCC 14579 / DSM 31).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=226900;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608415; PubMed=12721630;

RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,

RA Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,

RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,

RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,

RA Overbeek R., Kyrpides N.;

RT "Genome sequence of Bacillus cereus and comparative analysis with

RT Bacillus anthracis.";

RL Nature 423:87-91 (2003).

DR EMBL: AE017010; AAP10697.1; -.

DR GO: 0016491; F:oxidoreductase activity; IEA.

DR GO: 0019164; F:pyruvate synthase activity; IEA.

KW Oxidoreductase; Pyruvate; Complete proteome.

SQ SEQUENCE 288 AA; 31424 MW; 725FBE0935FA8AD9 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 288;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6

Db 92 DGDGFA 97

# RESULT 9

ID Q98FP0

AC Q98FP0 PRELIMINARY; PRT; 310 AA.

DT 01-OCT-2001 (TReMBLrel. 18, Created)

DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Aromatic metabolite ABC transporter, YodE.

GN MLJ3683.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Phyllobacteriaceae; Mesorhizobium.

```

OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Takeuchi Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50527.1; -.
DR InterPro; IPR004360; Gly_bleo_diox.
DR Pfam; PF00903; Glyoxalase; 1.
KW Complete proteome.
SQ SEQUENCE 310 AA; 34199 MW; 2345BEB8C59FF390 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 123 DGDGFA 128

RESULT 10
Q92V50 PRELIMINARY; PRT; 310 AA.
ID Q92V50;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Gouzy J.,
RA Vorhoelter F.J., Hernandez-Lucas I., Becker A., Cowie A.,
RA Golding B., Fuehrer A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49267.1; -.
DE PIR; C95950; C95950.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR004360; Gly_bleo_diox.
DR Pfam; PF00903; Glyoxalase; 2.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 310 AA; 34289 MW; 0DAC969D8DCD3ADE CRC64;

Query Match 100.0%; Score 34; DB 16; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 123 DGDGFA 128

RESULT 11
Q9X7X2 PRELIMINARY; PRT; 310 AA.
ID Q9X7X2
AC Q9X7X2;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative beta-lactamase.
GN SC06777 OR SC6A5-26C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harter D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Squares S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939129; CAB39710.1; -.
DR PIR; T35417; T35417.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001279; Bactinase-like.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00753; lactamase B; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Complete proteome.
SQ SEQUENCE 310 AA; 34781 MW; EE6D457AA08171FC CRC64;

Query Match 100.0%; Score 34; DB 16; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 295 DGDGFA 300

RESULT 12
Q8UCN0 PRELIMINARY; PRT; 325 AA.
ID Q8UCN0;
AC Q8UCN0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Aromatic compounds dioxygenase.
GN ATU2453 OR AGR_C_4455.

```

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 CX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RMD 0509952;  
 RX MEDLINE=21156231; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutyavint T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nestor E.W.;  
 RA "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu P.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RA "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328 (2001).  
 DR EMBL; AE009193; AAL43441.1; ALT\_INIT.  
 DR EMBL; AE008159; AAK88190.1; -.  
 DR PIR; AC2878; AC2878.  
 DR PIR; E97654; E97654.  
 DR GO; GO:0016702; F:oxidoreductase activity, acting on single d. . .; IEA.  
 DR InterPro; IPR004360; Gly\_bleo\_diox.  
 DR Pfam; PF00903; Glyoxalase; 1.  
 DR Dioxigenase; Complete proteome.  
 SQ SEQUENCE 325 AA; 35533 MW; 3F516706C842152C CRC64;  
 Query Match 100.0%; Score 34; DB 16; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 DB 134 DGDGFA 139  
 PRELIMINARY; PRT; 326 AA.  
 RESULT 13  
 Q8XD41  
 ID Q8XD41  
 AC Q8XD41  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DE Orf, hypothetical protein.  
 GN YGFZ OR Z4236 OR ECS3770.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 CX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / BDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 DR EMBL; AE005520; AAG58026.1; -.  
 DR EMBL; AP002563; BAB37193.1; -.  
 DR PIR; B91100; B91100.  
 DR PIR; F85945; F85945.  
 DR GO; GO:0004047; F:aminomethyltransferase activity; IEA.  
 DR InterPro; IPR006222; GCV\_T.  
 DR Pfam; PF01571; GCV\_T; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 326 AA; 36058 MW; 14F326302BEEF55A CRC64;  
 Query Match 100.0%; Score 34; DB 16; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 DB 78 DGDGFA 83  
 PRELIMINARY; PRT; 326 AA.  
 RESULT 14  
 Q8FE70  
 ID Q8FE70  
 AC Q8FE70  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN YGFZ OR C3479.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 CX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=06:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AE016766; AAN81927.1; -.  
 DR GO; GO:0004047; F:aminomethyltransferase activity; IEA.  
 DR InterPro; IPR006222; GCV\_T.  
 DR Pfam; PF01571; GCV\_T; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 326 AA; 36258 MW; 32D6E4C3CCBCB60F CRC64;  
 Query Match 100.0%; Score 34; DB 16; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGDGFA 6  
 DB 78 DGDGFA 83

```
RESULT 15
Q9HPA2 PRELIMINARY; PRT; 536 AA.
AC Q9HPA2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DT Htr17 transducer.
GN HTR17 OR VNG1733G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2050483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.,
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005078; AAG19968.1; -.
DR PIR; D84325; D84325.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0004871; F.signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Mc.chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 536 AA; 56815 MW; 3B33D2FD8701EB04 CRC64;

Query Match 100.0%; Score 34; DB 17; Length 536;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGDGFA 6
Db 358 DGDGFA 363
```

Search completed: June 1, 2004, 18:09:54  
Job time : 41 secs